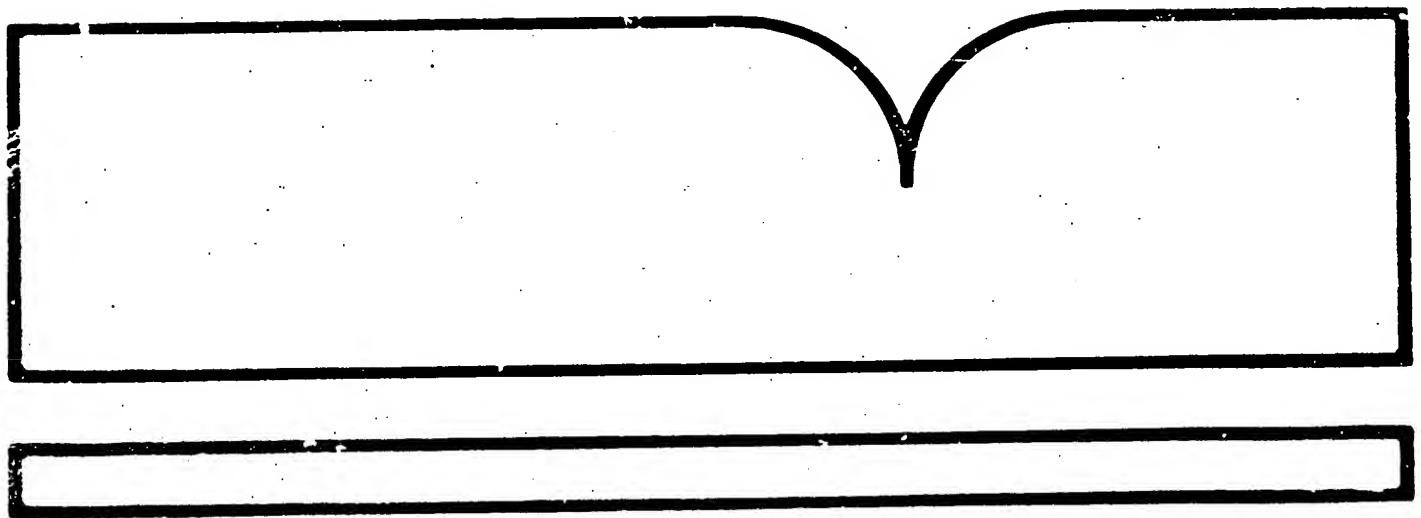


PB89-186233

Nucleotide and Amino Acid Sequences of the
Four Variable Domains of the Major Outer
Membrane Proteins of 'Chlamydia trachomatis'

(U.S.) Department of Health and Human Services
Washington, DC

17 Mar 89



BIBLIOGRAPHIC INFORMATION

PB89 186233

Report Nos: PAT-APPL-7-324 664

Title: Nucleotide and Amino Acid Sequences of the Four Variable Domains of the Major Outer Membrane Proteins of 'Chlamydia trachomatis'.

Date: Filed 17 Mar 89

Authors: H. D. Caldwell, Y. Ying, Y. X. Zhang, and N. G. Watkins.

Performing Organization: Department of Health and Human Services, Washington, DC.

Type of Report and Period Covered: Patent Application.

NTIS Field/Group Codes: 57E, 57F, 57D, 90D

Price: PC A04/MF A01

Availability: Available from the National Technical Information Service,
Springfield, VA. 22161

Number of Pages: 52p

Keywords: *Chlamydia trachomatis, *Urogenital diseases, *Patent applications, Antigen antibody reactions, Clinical medicine, Serology, *Diagnostic techniques, *Immunoassay, Amino acid sequence.

Abstract: Based on the immunological and biological relationships among Major Outer Membrane Proteins Variable Domains (MOMP VD) structure and function, and the fact that the MOMP VDs appear to be the major variable sequences between serovars, the present inventors identified the nucleotide and amino acid sequences of the MOMP VDs I-IV of *C. trachomatis* serovars Ba, D, E, F, G, H, I, J, K, and L3, and confirmed those of serovars B, A, C, Ll and L2. The present invention will allow the production of reagents and methodologies applicable in the development of new diagnostic tests for *C. trachomatis* infections and serological tests for serotyping. Specifically, species or sero-type specific oligopeptides conjugated to an array of reporter or detector groups could be used in various conventional non-immunologic assays including standard hybridization techniques.

APPLICATION FOR UNITED STATES PATENT

HHS

Inventor(s):

Harlan D. CALDWELL
Yuan YING
You-Xun ZHANG
Nancy G. WATKINS

Invention:

NUCLEOTIDE AND AMINO ACID SEQUENCES
OF THE FOUR VARIABLE DOMAINS OF THE
MAJOR OUTER MEMBRANE PROTEINS OF
CHLAMYDIA TRACHOMATIS

Cushman, Darby & Cushman
1615 L Street, N.W.
Eleventh Floor
Washington, D.C. 20036-5601
Attorneys
Telephone: 861-3000

SPECIFICATION

BACKGROUND OF THE INVENTION

In the United States, urogenital infections with Chlamydia trachomatis are the leading sexually transmitted disease with an estimated 10 million new cases reported each year. The urogenital infections result in involuntary infertility in 100-200,000 women each year. It is estimated that equal or greater numbers of these infections occur in European countries. In addition, ocular infection with C. trachomatis results in blinding trachoma which afflicts approximately 500 million individuals from developing countries. Improved diagnostic methodologies and vaccines or immunoprophylactic preventative measures are need to manage and control the disease.

Chlamydia trachomatis isolates occur as 15 distinct serovars. Based on serological relatedness, these 15 serovars have been divided into three serogroups; B serogroup (serovars B, Ba, D, E, L1 and L2), intermediate serogroup (serovars F, G, K and L3), and C serogroup (serovars A, C, H, I and J). The antigen that confers serovar and serogroup-specificities to chlamydia is the major outer membrane protein (MOMP), and protective immunity developed during chlamydial ocular infection is thought to be directed at serovar and serogroup MOMP determinants.

The genes encoding the MOMP of the C. trachomatis serovars A, B, C, L1 and L2 have been cloned and sequenced (Pickett et al., FEMS Microbiol. Lett. 42:185-190 (1987); Stephen et al., J. Bacteriol. 168:1277-1282 (1986); and Stephens et al., J. Bacteriol. 162:3879-3885 (1987); Baehr et al., Proc. Natl. Acad. Sci. USA, 85:4000-4004

(1988)). Comparative analysis of their amino acid sequences show the MOMP genes to be highly conserved structures that contain four evenly spaced domains whose sequences vary among the different serovars.

5 The locations of these nucleotide and amino acid sequence variable domains (VDs) are:

- VDI - nucleotides 256-315 and residues 64-83;
- VDII - nucleotides 481-546 and residues 139-160;
- VDIII - nucleotides 736-777 and residues 224-237;
- 10 VDIV - nucleotides 928-1017 and residues 288-317.

Epitope mapping has shown that three of the four VD domains (I, II and IV) contain contiguous antigenic determinants which elicit the formation of serovar, subspecies (determinants common to three or
15 more serovars within a serogroup), serogroup or species-specific antibodies.

Variable domains I and II, which demonstrate the greatest amount of inter serogroup sequence variation, are the locations of the serovar-specific
20 determinants. Variable domain IV is the largest of the MOMP VDs and is located near the C-terminus of the protein. VDIV is the location of subspecies and serogroup antigenic determinants, as well as a highly conserved species-specific antigenic
25 determinant. Antigenic determinants have not been mapped to VDIII, the smallest and least variable domain of the MOMP genes.

Variable domains I, II and IV of the MOMP of C. trachomatis protrude from the cell surface towards the external environment as demonstrated by their susceptibility to cleavage by trypsin, and
5 their accessibility to antibody binding. Trypsin cleavage in both VDII and VDIV, but not within VDIV alone, decreases chlamydial attachment to HeLa cells suggesting that these domains, or conformational
10 MOMP structures that are dependent on the integrity of these domains, may function as a chlamydial ligand.

Based on the immunological and biological relationships among MOMP VD structure and function, and the fact that the MOMP VDs appear to be the
15 major variable sequences between serovars, the present inventors sought to identify the nucleotide and amino acid sequences of the MOMP VDs I-IV of C. trachomatis serovars Ba, D, E, F, G, H, I, J, K and L3, and to confirm those of serovars B, A, C, L1 and
20 L2.

SUMMARY OF THE INVENTION

The amino acid sequences of the major outer membrane protein genes from Chlamydia trachomatis serovars A, B, C, L1 and L2 have been shown to be
25 conserved with the exception of four variable

domains in which major neutralizing and serotyping antigenic determinants have been located. Using oligonucleotide primer extension sequencing of MOMP mRNA, the nucleotide sequences of the four MOMP VDs of the remaining 10 *C. trachomatis* serovars were determined by the present inventors and the amino acid sequences deduced therefrom. The sequencing technique was rapid and required minimal amounts of total RNA, i.e., only 35 µg of RNA from chlamydial infected HeLa cells was needed to sequence all 4 variable domains for a given serovar. Use of the disclosed technique also circumvented the more standard and laborious approach of molecular cloning and direct DNA sequencing which had been employed in the past to determine the sequences of MOMP gene to serovars A, B, C, L1 and L2.

As noted above, the nucleotide sequences were employed to deduce the corresponding amino acid sequences. Comparative analysis of the amino acid sequence homology of the four variable domains permitted the present inventors to separate the fifteen serovars into three serogroups:

- Group 1 - serovars B, Ba, D, E, L1 and L2;
- Group 2 - serovars G and F;
- Group 3 - serovars A, C, H, I, J, K and L3.

The four variable domains were also analyzed for immunogenicity based on the hydrophilicity and

charge values of each domain (Baehr et al., supra).
The MOMP VDs with the greatest total hydrophilicity
and charge values were found to be the location of
antigenic determinants recognized by MOMP specific
5 monoclonal antibodies (Stephens et al., supra; Baehr
et al., supra).

The nucleotide, amino acid sequences and
hydrophilicity/charge value analyses are
advantageous because they will assist in the
10 selection of appropriate MOMP antigenic determinants
to be used in the construction of synthetic
peptides, subunits or recombinant chlamydial
vaccines. Examples of such recombinant vaccines
include infectious enteric vectors (avirulent
15 Salmonella typhimurium and/or Sabin Type I polio
virus) expressing MOMP variable domain sequences as
antigenic chimeras or hybrid molecules.

The present invention will allow the
production of reagents and methodologies applicable
20 in the development of new diagnostic tests for C.
trachomatis infections and serological tests for
serotyping. Specifically, species or sero-type
specific oligopeptides conjugated to an array of
reporter or detector groups could be used in various
25 conventional non-immunologic assays including
standard hybridization techniques.

Likewise, reagents for use in immunologic or serologic tests could be prepared based on the sequences presently disclosed; the conserved VD-IV peptide sequence TTLNTTIAG would permit species
5 specific identification, while the non-conserved sequences would enable one to make sero-type specific identification. It is contemplated by the present inventors that such reagents would be suitable for use in standard immunologic procedures
10 usch as ELISA and radioimmunoassay techniques.

The invention will be more fully described in the detailed description and drawings which follow.

DESCRIPTION OF THE DRAWINGS

Fig. 1. Location of the MOMP gene sequences
15 to which synthetic complementary oligonucleotides were constructed and used for primer extension mRNA sequencing of MOMP VDs. The MOMP gene sequences identified by single lined boxes are those to which synthetic oligonucleotides were constructed. The
20 double lined boxes identify the four variable domain regions of B, L2, A, C, and F MOMP genes.

Fig. 2. Sequencing of serovar B and A MOMP
VDII by primer extension of MOMP mRNA. Total RNA
was isolated from HeLa 229 cells infected with
25 serovars B and A and annealed to kinased

oligonucleotides BFP-2 and CP-2 (Table 1),
respectively. The oligonucleotides were
complementary to the non-coding strand of MOMP gene
sequences 50 nucleotides downstream 3' of VDII. The
5 RNA templates were transcribed with reverse
transcriptase, and the sequences of the transcribed
DNAs determined by the dideoxynucleotide chain
termination method. Autoradiographs were read to
depict + strand MOMP DNA sequences. The nucleotide
10 numbers shown for each MOMP are the beginning and
end of the sequence encoding VDII of each protein.
The results shown are representative of sequencing
results obtained for all other serovars' VDs.

Fig. 3. (A, B, C and D) The nucleotide
15 sequences of MOMP VDs I, II, III and IV,
respectively, of the 15 *C. trachomatis* serovars.
The boxed regions of the sequences identify each of
the four VDs. Serovars B, F and C MOMP sequences
were used as prototypes for comparative sequence
20 analysis. The nucleotide positions shown are for
serovar B MOMP. Exact nucleotide positions for
other MOMP VD sequences differ slightly because of
insertions and/or deletions within the VDs.

Fig. 4. (A, B, C and D) The deduced amino
25 acid sequences of the MOMP VDs 1, 2, 3 and 4,
respectively, of the 15 *C. trachomatis* serovars.

The boxed regions of the sequences identify each of the four VDs. Serovars B, F and C MOMP amino acid sequences were used as prototypes for comparative sequence analysis.

5 Fig. 5. C. trachomatis serovar B MOMP
variable domain sequence.

Fig. 6. C. trachomatis serovar Ba MOMP
variable domain sequence.

10 Fig. 7. C. trachomatis serovar D MOMP
variable domain sequence.

Fig. 8. C. trachomatis serovar E MOMP
variable domain sequence.

Fig. 9. C. trachomatis serovar L1 MOMP
variable domain sequence.

15 Fig. 10. C. trachomatis serovar L2 MOMP
variable domain sequence.

Fig. 11. C. trachomatis serovar F MOMP
variable domain sequence.

20 Fig. 12. C. trachomatis serovar G MOMP
variable domain sequence.

Fig. 13. C. trachomatis serovar C MOMP
variable domain sequence.

Fig. 14. C. trachomatis serovar A MOMP
variable domain sequence.

25 Fig. 15. C. trachomatis serovar H MOMP
variable domain sequence.

Fig. 16. C. trachomatis serovar I MCMP
variable domain sequence.

Fig. 17. C. trachomatis serovar J MOMP
variable domain sequence.

5 Fig. 18. C. trachomatis serovar K MOMP
variable domain sequence.

Fig. 19. C. trachomatis serovar L3 MOMP
variable domain sequence.

DETAILED DESCRIPTION OF THE INVENTION

10 All references cited below are specifically
incorporated into the specification by reference.

I. MATERIALS AND METHODS

A. Chlamydiae.

C. trachomatis serovars: A/Har-13, B/TW-
15 5/OT, Ba/AP-2, C/TW-3/OT, D/UW-3/Cx, E/Bour, F/IC-
Cal-13, G/UW-57/Cx, H/UW-4/Cx, I/UW-12/Ur, J/UW-
36/Cx, K/UW-31/Cx, LGV/L1-440, LGV/L2-434 and
LGV/L3-404 were employed herein and were immunotyped
by the micro-immunofluorescence (micro-IF) procedure
20 of Wang et al. (J. Infect. Dis. 152:791-800 (1985)).
Serovars A, E and F were obtained from the American
Type Culture Collection (ATCC, Rockville, MD). All
other serovars were obtained from Dr. Cho-Chou Kuo,

Department of Pathobiology, University of
Washington, Seattle, WA.

B. RNA extraction.

Approximately 2×10^8 HeLa 229 cells grown in
5 stationary culture were each infected with one
chlamydial serovar. Infected HeLa cells were
removed from the culture flasks with trypsin 24-
hours post-infection for serovars L1, L2 and L3, and
36-hours post-infection for all other serovars.
10 Cell suspensions were then pooled, centrifuged and
washed twice in cold (4°C) phosphate-buffered saline
(PBS), 0.02 M sodium phosphate, 0.15 M NaCl, pH 7.2.
Cells were lysed in 5 ml of cold guanidine
isothiocyanate lysis buffer (Current Protocols in
15 Molecular Biology, Wiley Inter Science, p. 4.2.4
(1987)). The resulting suspension was then gently
refluxed through an 18 gauge needle 15-20 times to
shear the DNA.

Total RNA was extracted (2x) from the
20 solution using hot acidic phenol (65°C, pH 5.0),
followed by two extractions with chloroform/isoamyl
alcohol (24:1 vol/vol). The RNA from each pool was
precipitated with two volumes of cold ethanol (-20°C).
The precipitate was evaporated to dryness using
25 a Speed-vac (Bachofar, Reutlingen, FRG). The RNAs

were resuspended in glass distilled water at a concentration of 20-40 µg/ml and stored at -70°C.

C. Synthetic oligonucleotides.

Oligonucleotides (21 mers) complementary to
5 the + strand MOMP DNAs flanking the 3' end of each
VD were synthesized using a SAM1 automated synthesis
instrument (Biosearch, San Rafael, CA) by the N-
methylimidazole phosphotriester method described by
Grayston et al. (J. Infect. Dis. 132:87-104 (1975))
10 and according to the manufacturer's instructions.
De-blocked oligonucleotides were evaporated to
dryness using a Speed-Vac, and were then dissolved
in distilled water at a concentration of 10-30 mg/ml
and stored at -20°C. Oligonucleotides were
15 constructed according to the published sequences of
the A, B, C, L2 MOMP gene serovars (Stephens et al.,
J. Bacteriol. 168, supra and Baehr et al., supra),
as well as the unpublished F serovar sequence which
was determined by the present inventors.

20 The MOMP gene sequences to which synthetic
complementary oligonucleotides were constructed, and
their location in the MOMP genes are shown in Fig.
1. The oligonucleotide number, its sequence, and
the MOMP serovar and VD for which the
25 oligonucleotide was used for primer extension mRNA
sequencing are summarized below in Table I.

oligonucleotide was used for primer extension mRNA sequencing are summarized below in Table I.

TABLE 1. Oligonucleotides Used in Primer-extension
mRNA Sequencing of *C. trachomatis* MOMP VDs

Oligonucleotide number	Oligonucleotide sequence	Serovars and their VDS sequencing with oligonucleotide primers
BP-1	5'-ATCCCAAATATTCAATGCCAT-3'	VDIs of B, Ba, D, E, G, L1, L2
CP-1	ATCCCAGATATTTAATGCCAT	VDIs of A, C, H, I, J, K, L3
FP-1	ATCCCAAATATTCAATGTCAT	VDIs of F, G
BFP-2	TCCACATTCCCACAAAGCTGC	VDIIs of B, Ba, D, E, F, G, L1, L2
CP-2	CCCACATTCCCAGAGAGCTGC	VDIIs of A, C, H, I, J, K, L3
BP-3	TCTGTAAGAGAGAGCTAAACT	VDIIIs of B, Ba, D, E, L1, L2
CP-3	AGTGAACATATTTAATCTGTA	VDIIIs of A, C, H, I, J, K, L3
FP-3	AGTGAACATATTGAGTCTGTA	VDIIIs of F, G
P-4-1	AATACCGCAAGATTTTCTAGA	VDIVs of A, B, Ba, C, D, E, F, H, I, J, K, L1, L2
P-4-2	TCCTACTGCAATACCGCAAGA	VDIVs of L3, G, (ND*)
P-4-3	TTTTCTAGATTTTCATCTTGTT	VDIV of J, (ND*)

D. Primer extension and DNA sequencing.

The dideoxynucleotide chain terminating method of Sanger et al. (Proc. Natl. Acad. Sci. USA 74:5463-5467 (1977) modified for the use with reverse transcriptase and RNA templates as described by Lane et al. (Proc. Natl. Acad. Sci. USA 82:6955-6959 (1985)) was used for sequence determinations. In summary, oligonucleotides were 5' end labeled with [³²P]dATP (3,000 Ci/mmol - New England Nuclear, Boston, MA) using T4 polynucleotide kinase (10,000 U/ml - New England Biolabs, Beverly, MA) as described by Maniatis et al. in A Laboratory manual. pp. 1-545 (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982). Labeled oligonucleotides were isolated on Sephadex G-25 columns (Boehringer Mannheim Biochemicals, Indianapolis, IN). The oligonucleotide primers were then hybridized to MOMP mRNA templates in 12.5 µl reaction mixtures containing 2.5 µl 5x annealing buffer (250 mM Tris-HCl, 300 mM NaCl, 250 mM DTT, and 5 mM EDTA, pH 8.3), 7.5 µl of total RNA (3-10 µg/µl). and 2.5 µl of kinased primer (2-2.5 µM).

The mixtures were heated at 90°C for 3 minutes, centrifuged and then allowed to cool slowly to 45°C over a 30 minute period. 2 µl of the hybridized mixture was then added to each of four tubes containing 3µl of reaction mixture (250 mM Tris-HCl, pH 8.3, 300 mM NaCl, 50 mM DTT, 30 mM MgCl₂, 2 mM dNTPs, 0.4 mM ddNTPs and 0.6 - 0.8 U/µl avian myeloblastosis virus reverse transcriptase (20,000 U/ml - Pharmacia Fine Chemicals, Inc., Piscataway, N.J.). The mixtures were incubated at 45°C for 40 minutes then the reactions were stopped by evaporating the tubes to dryness in a Speed-vac.

Following evaporation, the samples were resuspended in 3 μ l of 100% formamide containing 0.8% xylene cyanol and 0.8% bromophenol blue and heated at 95°C for 3 minutes. The samples were then promptly loaded on sequencing gels (8% polyacrylamide) to achieve separation. Following separation the gels were fixed with methanol, dried then autoradiographed.

II. Sequence Analysis

10 A. Primer extension and DNA sequencing.

Using the oligonucleotide primers described above, the VD-encoding portions of the MOMP's mRNA were sequenced. A representative sequencing gel is shown in Fig. 2 wherein MOMP B-VDII and A-VDII are shown. Oligonucleotides BFP-2 and CP-2 were employed in this sequencing as primers. The mRNA derived MOMP sequences depicted are identical to the previously published genomic DNA sequences for the MOMP B-VDII and A-VDII (Stephens et al., supra and Baehr et al., supra), thereby demonstrating both the feasibility and accuracy of this sequencing method.

B. Sequence analysis.

The synthetic oligonucleotides used for sequencing the MOMP VDs of the 15 C. trachomatis serovars are those depicted previously in Table 1. The location of the complementary sequences in the MOMP genes of serovars B, L2, A, C and F are shown in Fig. 3. All oligonucleotides were 21 mers and were complementary to conserved nucleotide sequences 40 to 60 nucleotides downstream (3') from the VD to be sequenced. Because inter-serogroup (but not intra-serogroup) nucleotide variation occurs within these regions, oligonucleotides complementary to

serogroup specific MOMP sequences were constructed for use in the sequencing of the variable domains of the serovars within serogroup. The 40 to 60 nucleotide distance between the VD and the downstream oligonucleotide primer site was necessary to produce high resolution sequences of the VDs.

The nucleotide and deduced amino acid sequences of the 4 variable domains for each of the 15 *C. trachomatis* serovars are individually depicted in Figs. 5-19. These sequences are also presented comparatively in Fig. 3 (nucleotide sequences) and Fig. 4 (deduced amino acid sequences). The B, F and C serovars' MOMP VDs are used as prototype sequences for comparative analyses of serovar VDs within the B, intermediate and C serogroup, respectively. A discussion of the results of the comparative analyses of the nucleotide and amino acid sequences of the four VDs for each *C. trachomatis* serovar is presented below.

1. Variable Domain I. VDI contains 60 nucleotides encoding 20 amino acids for all B serogroup serovars, 66 nucleotides encoding 22 amino acids for serovars K, L3 and all C serogroup serovars, and 63 nucleotides encoding 21 amino acids for serovars F and G. The nucleotide sequences of VDIs from B serogroup serovars are highly conserved with only 3 to 5 substitutions per serovar as compared to the B serovar prototype sequence. The deduced amino acid sequences of VDIs of B serogroup serovars are also conserved with only 2 to 4 substitutions per serovar, with the majority of substitutions occurring within a 12 amino acid region (residues 68-79) that resides in the central portion of the domain.

In contrast to the above, serovars in the C serogroup and serovars K and L3 show greater nucleotide variation in VDI than the B serogroup. Compared to C VDI, the other C serogroup serovars have 1 to 9 nucleotide substitutions in VDI that result in 0 to 4 amino acid changes. As above, the majority of these substitutions occur in the central region of the domain.

VDIs of serovars F and G are identical in their nucleotide and amino acid sequences.

2. Variable Domain II. VDII contains 63 nucleotides encoding 21 amino acids in the MOMP of serovars D, E, L1, F and G. The VDIIs of the remaining 10 serovars have 66 nucleotides encoding 22 amino acids. VDII shows more sequence variation than VDI in all serovars. As compared to serovar B, other B serogroup serovars have 2 to 19 nucleotide substitutions resulting in 2 to 8 amino acid changes. The least amount of variation occurs between serovars B and Ba in which there are two nucleotide substitutions that each result in an amino acid substitution. Most of the amino acid substitutions occur within an 8 amino acid region in the central part of the domain (residues 144 to 151).

The VDIIs of C serogroup serovars show considerable sequence variation having 4 to 8 nucleotide substitutions which result in 3 to 6 amino acid changes. A similar amount of variation was found in VDII sequences between the F and G serovars of the intermediate serogroup. Compared to serovar F VDII sequences, serovar G VDII has 6 nucleotide substitutions resulting in 3 amino acid changes.

3. Variable domain III. VDIII is the smallest and least variable of the four domains. VDIII contains 42 nucleotides which encode 14 amino acids. The VDIII sequences of A, H, J and L3 are identical to each other. Serovars D, L1 and F have identical sequences which differ from serovar B by 3 amino acid substitutions. The remaining serovars have sequences differing from serovar B by 1 to 4 amino acid substitutions.

4. Variable Domain IV. VDIV is the largest variable domain containing 96 nucleotides encoding 32 amino acids in serovars A, I and all B serogroup serovars, and 99 nucleotides encoding 33 amino acids in the remaining serovars. The VDIV domain can be separated into three distinct regions based on sequence homology: (i) the N-terminal region (residues 288-295), (ii) the central region (residues 296-306), and (iii) the C-terminal regions (residues 307-317).

Using micro-immunofluorescence polyclonal mouse sera, the 15 serovars have been separated into three serogroups. Group 1 or the B serogroup contains serovars B, Ba, D, E, L1 and L2; group 2, the intermediate serogroup, serovars F and G; and group 3, the C serogroup serovars A, C, H, I, J, K and L3.

Within the B serogroup, serovars B and Ba have identical sequences in the N-terminal regions, as do serovars D and E. There are three amino acid substitutions between serovars B-Ba and D-E. Serovars L1 and L2 contain one and two additional amino acid substitutions in this region, respectively. Within the C serogroup, serovars C,

H, I, J and K have identical sequences in the N-terminal region of VDIV. Serovar L3 contains a single amino acid substitution. Serovar A contains the most variable sequence in this region, containing 4 sequence substitutions as compared to the prototype C serovar sequence. Serovars F and G differ in this region by two amino acids.

The central region of VDIV is the most highly conserved sequence among the VDs of the chlamydial serovars. A nine amino acid sequence (TTLNPTIAG) is conserved in all serovars except serovar K, which contains a threonine (T) instead of an alanine (A) (residue 303). The central NPT sequence in VDIV is present in all 15 serovars.

The C-terminal region of VDIV is highly conserved among the B serogroup serovars. In contrast, this region of VDIV in C serogroup serovars shows a significant amount of amino acid variation, although the amino acid composition is similar (predominately S, T, E and A). Serovars F and G differ by three amino acid residues in this region.

In contrast to the historical classification noted above, wherein serovars K and L3 are placed within the C serogroup, immunotyping with monoclonal antibodies specific to MOMP has been recently employed and serovars K and L3 have been reclassified to the intermediate serogroup. Based on the present invention and comparative MOMP VD amino acid homology analysis, however, serovars K and L3 clearly belong in the historical C serogroup.

C. Inter-and intra-serogroup amino acid homologies.

The percent of amino acid homology and the total number of amino acid substitutions for the VDs of each serovar are shown in Table 2.

5

TABLE 2. Comparative amino acid homologies of MOMP VOs of the 15 *E. trachomatis* serovars^a

B	B ^a	D	E	L1	L2	F	G	C	A	H	I	J	K	L3
B	100 ^b 94.2 (5)	80.2 (17)	76.7 (20)	77.9 (19)	77.9 (19)	43.7 (49)	42.5 (50)	38.2 (55)	40.9 (52)	40.4 (53)	43.2 (50)	42.7 (51)	42.7 (51)	41.6 (52)
B ^a	100 (0)	77.9 (19)	76.7 (20)	76.7 (20)	76.7 (20)	43.7 (49)	43.7 (49)	37.1 (56)	39.8 (53)	39.3 (54)	52 (51)	41.6 (52)	41.6 (52)	40.4 (53)
D	100 (0)	100 (0)	82.4 (15)	85.9 (12)	76.7 (20)	44.8 (48)	43.7 (49)	38.2 (55)	42.7 (51)	39.3 (54)	42 (51)	41.6 (52)	40.4 (53)	41.6 (52)
E	100 (0)	100 (0)	100 (0)	82.4 (15)	76 (21)	39.1 (53)	43.7 (49)	40.4 (53)	43.2 (50)	40.2 (53)	42 (51)	43.8 (50)	41.6 (52)	43.8 (52)
L1		100 (0)		100 (0)	77.3 (20)	44.8 (48)	44.8 (48)	40.4 (53)	43.2 (50)	42.7 (51)	44.3 (49)	46.1 (48)	41.6 (52)	42.7 (51)
L2		100 (0)		100 (0)	100 (0)	40.2 (52)	42.2 (50)	36 (57)	39.8 (53)	39.3 (54)	42 (51)	40.4 (53)	38.2 (55)	43.8 (52)
F						100 (0)	85.1 (13)	34.8 (58)	37.5 (55)	34 (57)	34.1 (58)	37.1 (56)	32.5 (60)	39.3 (54)
G						100 (0)	100 (0)	39.3 (54)	40.9 (52)	36.2 (55)	37.5 (55)	40.4 (53)	36 (57)	43.8 (50)
C								100 (0)	77.5 (20)	83.1 (15)	86.4 (12)	88.8 (12)	85.4 (10)	85.4 (13)
A									100 (0)	76.2 (23)	81.8 (16)	84.3 (16)	75.3 (22)	79.9 (19)
H										100 (0)	83 (15)	84.5 (12)	83.1 (15)	83.1 (15)
I											100 (0)	87.6 (11)	81.8 (16)	86.4 (12)
J												100 (0)	86.5 (12)	87.6 (11)
K													100 (0)	80.9 (17)
L3														100 (0)

^aAmino acid sequences of the four VOs were used for calculating homologies.
^bPercent homology.
^cNumber of amino acid substitutions.

Inter-group (1, 2 and 3) VD amino acid homology is between 33 and 46%. Intra-group VD amino acid homology was between 74 and 94% for groups 1 and 3, and 85% for group 2. Therefore, classification of C. trachomatis isolates based on MOMP VD amino acid homology correlates extremely well with the historical serological classification and generally with immunotyping using MOMP specific monoclonal antibodies (see e.g., Wang et al., supra). The present inventors note, however, that recent immunotyping using monoclonal antibodies places serovars K and L3 within the intermediate serogroup (serovars F and G), whereas VD amino acid homology comparison shows serovars K and L3 to be more closely related to serovars within the C serogroup.

III. Immunogenicity based on hydrophilicity and charge values.

The VDs are hydrophilic and charged; properties indicating that these sequences are associated with the chlamydial cell surface and are potential antigenic sites. In fact, major antigenic sites which elicit the formation of protective serotyping antibodies have been mapped to VDI, VDII, and VIDIV of the MOMP of serovars A, B, C and L2 (Baehr et al., supra). However, the antigenic properties of the VDs of the remaining 11 serovars have not as yet been described.

In order to define the antigenicity of these domains, the MOMP VDs of the chlamydial serovars were analyzed for hydrophilicity and charge using standard measurements and calculations. The number of charged amino acids and the hydrophilicity value of sequences within each MOMP VD for the 15

serovars was determined. These values are provided in Table 3.

TABLE 2.
TABLE 3. The hydrophilicity values and charges of *G. trachomatis* MOMP VDs

VDI		VDII		VDIII		VDIV		VDV		C' end of VDIV								
Basic	Acidic	Basic	Acidic	Basic	Acidic	Basic	Acidic	Basic	Acidic	Basic	Acidic							
B 2	1	+1.8	1	2	+1.8	1	3	+4.3	1	2	+2.6	0	0	-5.1	1	3	+7.6	
Ba 2	1	+2.8	1	2	+2.2	1	3	+5.1	1	2	+2.6	0	0	-5.1	1	3	+7.6	
D 2	2	+8.8	2	4	+9.7	1	3	+3.6	1	1	+0.2	0	0	-5.1	1	3	+7.3	
E 2	2	+6.4	1	3	+7.8	0	2	-3.8	1	1	+0.2	0	0	-5.1	1	3	+7.5	
L1 2	1	+3.5	2	4	+12.5	1	3	+3.6	1	1	-1.9	0	0	-5.1	1	3	+7.9	
L2 2	1	+3.5	1	4	+7.7	1	3	+3.7	1	1	-0.5	0	0	-5.1	1	3	+7.5	
AVG ^d	1.3	+4.47	1.35	3.2	+6.95	0.8	2.8	+2.75	1	1.3	+0.53	0	0	-5.1	1	3	+7.57	
F 2	2	+3.5	1	2	+2.1	1	3	+3.6	1	2	-2.5	0	0	-5.1	0	2	+1.3	
G 2	2	+3.5	0	2	0	0	2	-2.8	2	1	+1.9	0	0	-5.1	0	2	+0.6	
AVG ²	2	+3.5	0.5	2	+1.05	0.5	2.5	+4.4	1.5	1.5	-0.3	0	0	-5.1	0	2	+0.95	
C 1	2	+1.5	2	1	-0.3	0	2	-2.7	1	2	+1.1	0	0	-5.1	1	3	+6.6	
A 2	3	+5.9	1	1	-3.1	0	3	+0.1	2	1	+3.0	0	0	-5.1	1	3	+6.6	
H 2	3	+8.8	4	1	+5.7	0	3	+0.1	1	2	+1.1	0	0	-5.1	1	3	+6.6	
I 2	3	+7.0	2	0	-0.1	0	3	-1.2	1	2	+1.1	0	0	-5.1	1	3	+6.6	
J 1	2	+1.5	1	0	-7.4	0	3	+0.1	1	2	+1.1	0	0	-5.1	1	3	+6.6	
K 1	3	+6.5	2	2	+0.1	0	3	-0.9	1	2	+1.1	0	0	-5.1	1	3	+7.3	
L3 1	3	+6.2	2	0	-3.6	0	3	+0.1	1	2	+1.4	0	0	-5.1	1	3	+7.3	
AVG ^f	1.4	2.7	+5.3	2	0.7	-1.24	0	2.9	-0.63	1.1	1.9	+1.4	0	0	-5.1	1	3	+6.8

^aTotal number of histidines, arginines and lysines.
^bTotal number of aspartic acids and glutamic acids.
^cVD = hydrophilicity value assigned by Kopp and Woods, Proc. Natl. Acad. Sci. USA 78:3924-2828 (1981)
^dAverage charge and NV for the B serogroup.
^eAverage charge and NV for the intermediate serogroup.
^fAverage charge and NV for the C serogroup.

5 As the data demonstrates, the VDs have a broad range of hydrophilicity and charge. The VDI sequences of serovars in the B serogroup had an average hydrophilicity value of 4.47, and contained an average of 2 positive (basic) charges and 1.3 negative (acidic) charges, while the VDII sequences had an average hydrophilicity value of 6.95, and contained an average of 1.35 positive charges and 3.2 negative charges. Based on these results, the VDII sequences of serovars within the B serogroup are predicated to have the greater immunogenic potential between the two.

Indeed, as reported by Stephens et al. (J. Exp. Med., ¹⁵supra) epitope mapping studies using antisera prepared against synthetic peptides corresponding to MOMP VDs show that serovar-specific determinants are located in the VDII of the MOMP of serovars B, C and L2, and a highly conserved species-specific MOMP determinant is mapped to VDIV as depicted in Table 4.

TABLE 4. The charge and hydrophilicity value of epitopes on *C. trachomatis* MOMP mapped by MAbs

Monoclonal antibody	Sequence of epitope ^a	Amino Acids		Hydrophilicity value	Location of epitope
		Basic	Acidic		
A-20	D-V-A-G-L-E-K-D-P-V	1	3	+4.1	A-VDI
L21-45	D-N-E-N-H-A-T-V-S-D-S-K-L-V	2	3	+6.8	L2-VDII
B-B6	N-N-E-N-Q-T-K-V-S-N-G-A-F-V	1	1	+0.9	B-VDII
B-B5	P-T-I-A-G-A-G-D-V-K-T-S-A-E-G	1	2	+3.8	B-VDIV C-terminus
L21-10	T-T-L-N-P-T-I-A-G	0	0	-5.1	VDIV of all serovars except K

^aSequence of epitope as determined by lambda gt11 epitope mapping according to Baehr et al., 1983.

In contrast to the above, VDI sequences of C serogroup serovars MOMP were more hydrophilic and charged than VDII. The average hydrophilicity for VDI was 5.3 with 1.4 positive and 2.7 negative charges, while the hydrophilicity value for VDII was -1.24 with 2 positive and 0.7 negative charges. These data suggest that VDI sequences of C serovars have a greater immunogenic potential than VDII sequences. This conclusion is supported by the work of Baehr et al., supra, who, using a lambda-gt11 epitope mapping procedure showed that the serovar-specific MOMP epitope for serovar A and L2 mapped to VDI and VDII, respectively, whereas subspecies, serogroup and species-specific epitopes mapped to VDIV. Specifically, the epitope recognized by monoclonal antibody A-20, which is specific for serovar A MOMP has been mapped to VDI (Table 4).

As with the C serogroup, the MOMP VDI sequences of the intermediate serovars F and G were more hydrophilic and charged than the VDII sequences. Several serovars (D, E and H) contain both VDI and VDII sequences which are extremely hydrophilic and charged, suggesting that in MOMPs of these serovars, both domains may be immunogenic.

Although antigenic sites have not yet been mapped to VDIII, its hydrophilicity and charge values implicate this domain as a potential immunogenic site as well.

As noted previously, VDIV has been divided into three regions, (i) N-terminal (residues 284-295), central (residues 296-306) and C-terminal (residues 307-317). The

hydrophilicity and charge for each region has been independently calculated.

The VDIV domain shows a consistent pattern of hydrophilicity and charge in each region for MOMP's of all serovars. The N-terminal region is charged but is weakly hydrophilic; the central region is consistently uncharged and is hydrophobic; and the C-terminal region is both highly charged and hydrophilic, with the exception of serovars F and G which have only moderate charge and hydrophilicity.

The single exception in correlating hydrophilicity and charge with known MOMP antigenic determinants is the species-specific MOMP epitope recognized by monoclonal antibody L21-10. This epitope is located in the central region of VDIV (TTLNPTIAG) and is uncharged and strongly hydrophobic (-5.1).

From the epitope data, it appears that serovar-specific determinants are located in MOMP VDI, VDII or both, while common or serogroup-specific MOMP antigenic determinants are located in VDIV.

ABSTRACT OF INVENTION

The nucleotide and deduced amino acid sequences of the four variable domains of the major outer membrane proteins of the 15 serovars of Chlamydia trachomatis are disclosed together with sequence and immunogenic analysis of these domains.

Figure 1
Location of map Cane Sequences to which Synthetic Complementary Oligonucleotides were Constructed

[illegible]

Figure 2

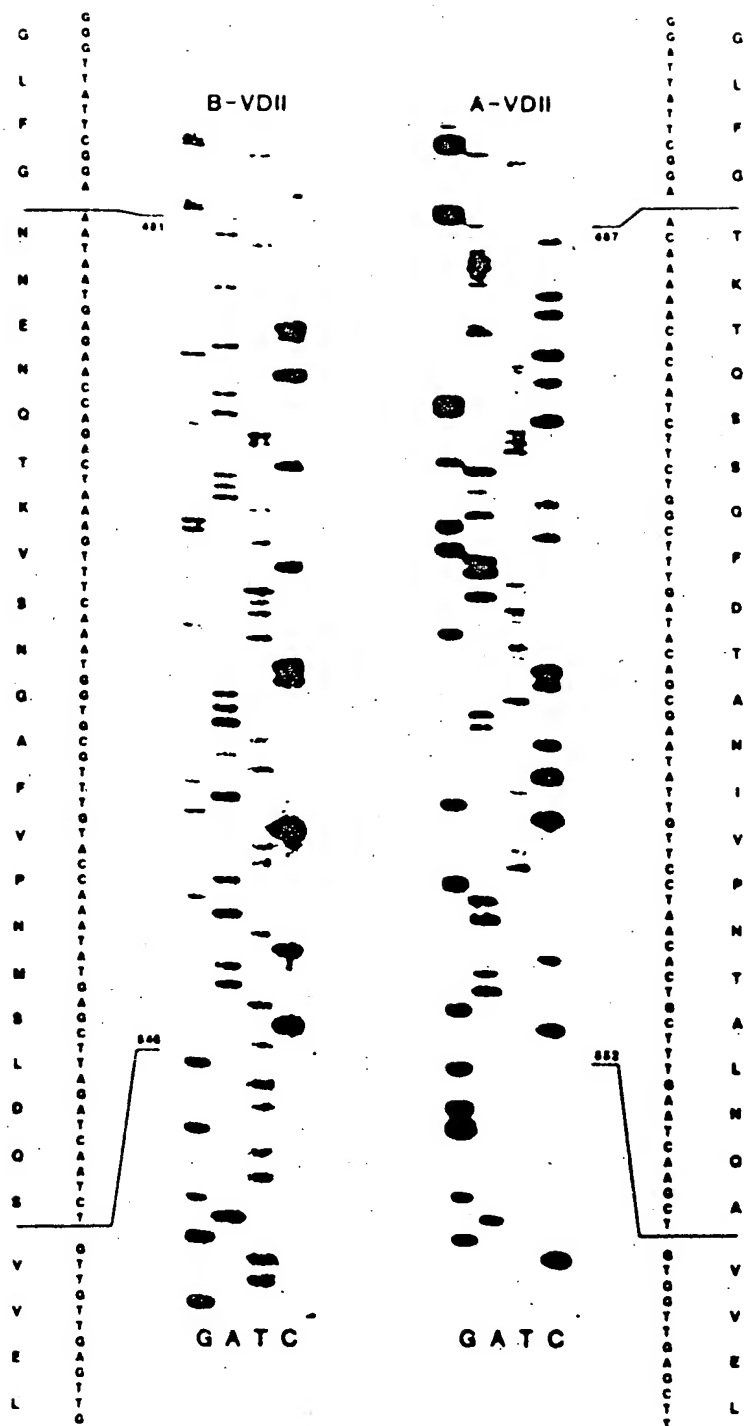


Figure 4

55	64	83	92
VDI			
DVNKEFCMG	-AKPTTTTGNVAPSTLT-ARE	NPAYGRHMQ	
A.....T.....		
D...SA.....		
	-D...S.....T...T.....		
A.....A.....C.....		
A.....A.....C.....		
DVNKEFEMG	EALAGASGNTTSTLSKL-VERT	NPAYGKHMQ	
DVNKEFCMG	AAPTTSADVAGLONPTTNVAP	NPAYGKHMQ	
EK..VA.....		
N.A.D.....K.....		
K.....E.....		
E.....		
	E.....T.....S.....		

132	139	160	166
VDII			
B	NLVGLFG	NNENQTKVSNAGAFVPMHSLDGS	WVELYT
B ₁	ST.....	
D		D....KT.YAES-.....F....	
E		D....ST.KTNS-.....	
L1		D....ST.KKD-.....F....	
L2		D...HAT...DSKL.....	
F	NLVGLFG	DGVNATKP-AADSIPNVQLNQS	WVELYT
G		...E...Q...T.....	
C	NLVGLFG	TKTQSSSFNTAKLIPNTALNEA	WVELYT
A	G.D..NIV.....Q.....	T
H		...K..D.....V..I...R.....	T
I	N.....V..A...Q.....	T
J	A.....N.F.....Q.....	T
K		...Y.K...N.V.....DR.....	T
L3		...TN.....V.....C.....	T

220	224	237	241
VDIII			
GYVG	KEFLDLTAGTDAA	TGTK	
S.....		
	...F.....		
	Q.F...A.I.....		
	...F.....		
	Q.F...K...GV.....		
GYVG	KEFLDLTAGTDAA	TGTK	
	Q...A.....		
GYVG	AEFLPLNITAGTEAA	TGTK	
D.....		
D.....		
D.I.....		
D.....		
	V...D.....		
D.....		

284	288	317	321
VDIV			
B	AQPK	SAETIFDVTTLNPTIAGAGGVKTS-EGGLR	DTMC
E _a	T.....	
D		...TA...T.....G.....	
E		...TA...T.....A.....	
L1		L.TA...T.....E..AN.....	
L2		...T.V.....A.....	
F	AQPR	LVTPVVDITTLNPTIAGGGSVAGANTEGGIS	DTMC
G	...K	...AK.....VA..S.....	
C	AQPK	LAEAILDVTTLNPTIAGKGSVVSACTDNELA	DTMC
A		...KPV...T.....T...SA-E.....	
H	T...AS.S..D.....	
I	T...SA-E.....	
J	T...AS.SE.D.....	
K	T...A...S.S.....	
L3		...V.....AS.SE.....	

Fig. 5. *C. trachomatis* serovar B major outer membrane protein variable domain sequences*

B-VD	256**	GGCCAAGCCTACAACACTACTACAGGCAATGCTGTAGCTCCATCCACTCTTACAGCAAGAGAG	315
		A K P T T T T G N A V A P S T L T A R E	
B-VDII	481	AATAATGAGAACCAGACTAAAGTTTCAAAATGGTGGTTTGACCAATATAGCTTAGATCAATCT	546
		N N E N Q T K V S N G A F V P N M S L D Q S	
B-VDIII	736	AAGGAGTTGCCCTCTTGATCTTACAGCAGGAACAGATGCTGCG	777
		K E L P L D L T A G T D A A	
B-VDIV	928	TCAGCCGAGACTATCTTTGATGTTACCACTCTGAACCCCAACTATTGCTGGAGCTGGCGATGTGAACACTAGCGCAGAGGGTCAGCTCGGA	1017
		S A E T I F D V T T L N P T I A G A G D V K T S A E G Q L G	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 7. *C. trachomatis* serovar D major outer membrane protein variable domain sequences*

D-VDI	256**	315
	GCCAAGCCTACAACTGATACAGGCAATAGTCAGCTCCACTCTTACAGCAAGAGAG	
	A K P T T D T G N S A A P S T L T A R E	
D-VDII	481	543
	GATAATGAAATCAAAAACGGTCAAGCGGAGTCTGTACCAATATGAGCTTTGATCAATCT	
	D N E N Q K T V K A E S V P N M S F D Q S	
D-VDIII	733	744
	AAGGAGTTTCCTCTTGATCTTACAGCAGGAACAGATGCTGCG	
	K E F P L D L T A G T D A A	
D-VDIV	925	
1014	TCAGCTACAGCTATTTTGTACTACACGCTTAACCCAACTATTGCTGGAGCTGGCGATGTGAAAACTGCCCCAGAGGGTCAGCTCGGA	
	S A T A I F D T T T L N P T I A G A G D V K T G A E G Q L G	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 8. *C. trachomatis* serovar E major outer membrane protein variable domain sequences*

256**	315
E-VDI GACAAGCCTACAAGTACTACAGGCAATGCTACAGCTCCAACCACTCTTACAGCAAGAGAG	
D K P T S T T G N A T A P T T L T A R E	
481	543
E-VDII GATAATGAAATCAAGCAGCGGTCAAAACGAATTCTGTACCAATATGAGCTTAGATCAATCT	
D N E N Q S T V K T N S V P N M S L D Q S	
733	744
E-VDIII CAAGAATTCCTCTTGCACTCATAGCAGGAAGTATGTCAGCG	
Q E F P L A L I A G T D A A	
E-VDIV	1014
925	
TCAGCTACAGCTATCTTTGATACCTACCAAGCTTAACCCAACTATTGCTGGAGCTGGCGATGTGAAAGCTACGAGGGTCAAGCTCGGA	
S A T A I F D T T T L N P T I A G A G D V K A S A E G Q L G	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 9. *C. trachomatis* serovar L1 major outer membrane protein variable domain sequences*

L1-VDI	256** GCCAAGCCTACAGCTACTACAGGCAATGCTGCAGCTCCATCCACTTGTACAGCAAGAGAG A K P T A T T G N A A A P S T C T A R E	315
L1-VDII	481 GATAATGAAATCAAGCACGGTCAAAAAGGATGCTGTACCAATATGAGCTTTGATCAATCT D N E N Q S T V K K D A V P N M S F D Q S	543
L1-VDIII	733 AAGGAGTTTCCTCTTGATCTTACAGCAGGAACAGATGCTGCG K E F P L D L T A G T D A A	774
L1-VDIV	925 TTAGCTACAGCTATCTTTGATACACCGCTTAACCCAACTATTGCTGGAGCTGGCGAGGTGAAGCTAACGCAGAGGCTCAGCTCGGA L A T A I F D T T T L N P T I A G A G E V K A N A E G Q L G	1014

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 10. *C. trachomatis* serovar L2 major outer membrane protein variable domain sequences*

L2-VDI	256** GCCAAGCCTACAACTGCTACAGGCAATGCTGCAGCTCCACTTGTACAGCAAGAGAG A K P T T A T G N A A A P S T C T A R E	315
L2-VDII	481 GATAATGAGAACCATGCTACAGTTTCAGATAGTAGCTTGACCAATATGAGCTTAGATCAATCT D N E N H A T V S D S K L V P N M S L D Q S	546
L2-VDIII	736 CAAGAATTCCTCTTGATCTTAAAGCAGGAACAGATGGTGTG Q E P L D L K A G T D G V	777
L2-VDIV	928 TCAGCTACAACTGTCTTTGATGTTACCACTCTGAACCCAACTATTGCTGGAGCTGGCGATGTGAAGCTAGCGCAGAGGGCTCAGCTCGGA S A T T V F D V T T L N P T I A G A G D V K A S A E G Q L G	1017

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 11. *C. trachomatis* serovar F major outer membrane protein variable domain sequences*

F-VDI	256	GAGGCTTAGCCGGAGCTTCTGGGAATACGACCTCTACTCTTTCAAAATTGGTAGAACGAACG	318
		E A L A G A S G N T T S T L S K L V E R T	
F-VDII	484	GATGGTGTAACGCCACGAAACCTGCTGCAGATAGTATTCTTAACGTGCAGTTAAATCAGTCT	546
		D G V N A T K P A A D S I P N V Q L N Q S	
F-VDIII	736	AAGGAGTTTCCTCTTGATCTTACAGCAGGAAACAGATGCAGCG	777
		K E F P L D L T A G T D A A	
F-VDIV	928		
	1020	TTGGTAACACCTGTTGTAGATATTACAACCCCTTAACCCCACTATTGCAGGATCGCGCAGTGTAGCTGAGCTAACACGGAAGACAGATATCT	546
		L V T P V V D I T T L N P T I A G C G S V A G A N T E G Q I S	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 12. *C. trachomatis* serovar G major outer membrane protein variable domain sequences*

G-VDI	256**	GAGGCTTTAGCCGGAGCTTCTGGGAATACGACCTCTACTCTTTTCAAAATTGGTAGAACGAACG	318
		E A L A G A S G N T T S T L S K L V E R T	
G-VDII	484	GATGCTGAAACGCCACGCGCTGCTGCAACAAGTATTCTTAACGTGCAGTTAAATCAGTCT	546
		D G E N A T Q P A A T S I P N V Q L N Q S	
G-VDIII	736	CAAGAATTCCCTCTTGCACTCACAGCAGGAAGTATGATGCAGCG	777
		Q E F P L A L T A G T L A A	
G-VDIV	928	TTGGCAAAACCTGTTGTAGATATTACAAACCCCTTAACCCAACTATTGCAGGATGGCGCAGTGTAGCTGCAGCTAACTCGGAAGGACAGATATCT	1020
		L A K P V V D I T T L N P T I A G C G S V A A N S E G Q I S	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 13. *C. trachomatis* serovar C major outer membrane protein variable domain sequences*

C-VDI	256**	321
	GCGGCGCCTACTACCGCGATGTAGCAGGCTTACAAAACGATCCAAACAACGTTGCTCGTCCA	
	A A P T T S D V A G L Q N D P T T N V A R P	
C-VDII	487	552
	ACAAAACACAATCTTCTAGCTTTAATACAGCGAATATTATTCCTAACACTGCTTTGAATGAAGCT	
	T K T Q S S S F N T A K L I P N T A L N E A	
C-VDIII	742	783
	GCGGAATTTCCACTTAATATTACCGCAGGAACAGAGCTGCG	
	A E F P L N I T A G T E A A	
C-VDIV	934	1026
	TTGGCTGAAGCAATCTTGGATGTCACTACTCTGAACCCGACTATCGCTGTAAAGGAAGTGTGCTCTGCGGAAACCGATAACGAACTGGCT	
	L A E A I L D V T T L N P T I A G K G S V S A G T D N E L A	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 14. *C. trachomatis* serovar A major outer membrane protein variable domain sequences*

A-VDI	256*	321
	GCGGGCCTACTACGCGGATGTAGCAGGCTTAGAAAACGATCCAGTAGCAATGTGTGCTGCCCCA	
	A A P T T S D V A G L E K D P V A N V A R P	
A-VDII	487	552
	ACAAAACACAATCTTCTGGCTTGTATACAGCGAATATTGTTCTAACACTGCTTTGAATCAAGCT	
	T K T Q S S G F D T A N I V P N T A L N Q A	
A-VDIII	742	783
	GCGGAATITCCACTTGATATTACCCAGGAACAGAGCTGCCG	
	A E F P L D I T A G T E A A	
A-VDIV	934	1023
	TTGGCTAAACCACTCTTGGATACCACTACTCTAAACCCGACCATCGCTGTAAGGAACCTGTGTCTCTTCGCAGAAAACGAACCTGGCT	
	L A K P V L D T T T L N P T I A G K G T V S S A E N E L A	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 15. *C. trachomatis* serovar H major outer membrane protein variable domain sequences*

	256**	321
H-VDI	GCGGCGCCTACTACCAACGATCGCAGCAGACTTACAACGATCCAAACAAATGTTGCTCGTCCA	
	A A P T T N D A A D L Q N D P K T N V A R P	
		552
	487	
H-VDII	ACAAAACAAATCTTCTGATTTTAATACAGCGAAGCTTGTTCTTAACATTTGTTGAATCGAGCT	
	T K T K S S D F N T A K L V P N I A L N R A	
		783
	742	
H-VDIII	GCGGAATTTCCACTTGATATTACCGCAGGAACAGAAGCTGCG	
	A E F P L D I T A G T E A A	
		1026
H-VDIV	TTGGCTGAAGCAATCTTGGATGTCACTACTCTAAACCCGACCATCGCTGTGTAAGGAACCTGTGGTCGGGAAGCGATAACGACCTGGCT	
	L A E A I L D V T T L N P T I A G K G T V A S G S D N D L A	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 16. *C. trachomatis* serovar I major outer membrane protein variable domain sequences*

I-VDI	256**	321
	GCGGGCGCCTACTACCAAGGATGTAGCAGGCTTAGAAAACGATCCAAACACAAATGTTGCTCGTCCA	
	A A P T T K D V A G L E N D P T T N V A R P	
I-VDII	487	552
	ACAIAACACAAATCTTCTTAATTAACAGCGAAGCTTGTTCTTAACGGCTGTTTGAATCAAGCT	
	T K T Q S S N F N T A K L V P N A A L N Q A	
I-VDIII	742	783
	GCGGAATTTCCACTTGATATATATCGCAGGAACAGAGCTGCG	
	A E F P L D I I A G T E A A	
I-VDIV	934	1023
	TTGGCTGAAGCAATCTTGGATGTCACTACTCTAAACCCGACCATCGCTGGTAAGGAAGTGTGGTCTCTTCCGCAGAAAACGAACCTGGCT	
	L A E A I L D V T T L N P T I A G K G T V S S A E N E L A	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

Fig. 17. *C. trachomatis* serovar J major outer membrane protein variable domain sequences*

-VDI	256**	321
	CGGGGGCTACTACGAGGATGTAGCAGGCTTACAAACGATCCAAACAATGTGCTCGTCCA	
	A A P T T S D V A G L Q N D P T T N V A R P	
-VDII	487	552
	ACAAACAACAAGCTTCTAGCTTTAATACAGCGAATCTTTTCCTAACACTGCTTTGAATCAAGCT	
	T K T Q A S S F N T A N L F P N T A L N Q A	
-VDIII	742	783
	GCGGAATTCACCTTGATATTACCGCAGGACAGAAAGCTGCG	
	A E F P L D I T A G T E A A	
-VDIV	34	1026
	TTGGCTGAAGCAATCTGGATGTCACCTACTCTAAACCCGACCATCGCTGTAAGGAACCTGTGGTCCGTAAGCGAAACGACCTGGCT	
	L A E A I L D V T T L N P T I A G K G T V V A S G S E N D L A	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

*Location of nucleotide sequence within the MOMP gene.

Fig. 18. *C. trachomatis* serovar K major outer membrane protein variable domain sequences*

K-VDI	256**	321
	GCGGGCCCTACTACCAGCGATGTAGAAGGCTTACAAAACGATCCAAACAACAAATGTTGGCTCGTCCA	
	A A P T T S D V E G L Q N D P T T N V A R P	
K-VDII	487	552
	ACAAAACACAATATTCTAAGTTTAATACAGCGAATCTTGTTCCTAACACTGCTTTGGATCGAGCT	
	T K T Q Y S K F N T A N L V P N T A L D R A	
K-VDIII	742	783
	GTGGAAITTCCTGATATTACCGCAGGAACAGAGCTGCG	
	V E F P L D I T A G T E A A	
K-VDIV	934	1026
	TTGGCTGAAGCAATCTTGGATGTCACCTACTCTAAACCGACCATCACTGTTAAGGAGCTGTGCTCTTCCGGAAGCGATAACGAACTGGCT	
	L A E A I L D V T T L N P T I T G K G A V V S G S D N E L A	

*Upper line for each VD is the nucleotide sequence, lower line is the deduced amino acid sequence using the one letter amino acid code.

**Location of nucleotide sequence within the MOMP gene.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)